A Channel that Heats Up

Tobias Koch Amos Lapidoth
ETH Zurich, Switzerland
Email: {tkoch, lapidoth}@isi.ee.ethz.ch

Paul P. Sotiriadis
Johns Hopkins University, Baltimore, USA
Email: pps@jhu.edu

***Abstract*— Motivated by on-chip communication, a channel model is proposed where the variance of the additive noise depends on the weighted sum of the past channel input powers. For this channel, an expression for the capacity per unit cost is derived, and it is shown that the expression holds also in the presence of feedback.**

I. Introduction

Continuous advancement in VLSI technologies has resulted in extremely small transistor sizes and highly complex microprocessors. However, on-chip interconnects responsible for on-chip communication have been improved only moderately. This leads to the "paradox" that local information processing is done very efficiently, but communicating information between on-chip units is a major challenge.

This work focuses on an emergent issue expected to challenge circuit development in future technologies. Information communication and processing is associated with energy dissipation into heat which raises the temperature of the transmitter/receiver or processing devices; moreover, the intrinsic device noise level depends strongly and increasingly on the temperature. Therefore, the total physical structure can be modeled as a communication channel whose noise level is data dependent. We describe this mathematically in the following subsection.

A. Channel Model

We consider the communication system depicted in Figure 1. The message M to be transmitted over the channel is assumed to be uniformly distributed over the set $\mathcal{M} = \{1, \ldots, |\mathcal{M}|\}$ for some positive integer $|\mathcal{M}|$. The encoder maps the message to the length-n sequence $X_1, \ldots, X_n$, where n is called the *block-length*. Thus, in the absence of feedback, the sequence X_1^n is a function of the message M, i.e., $X_1^n = \phi_n(M)$ for some mapping $\phi_n : \mathcal{M} \to \mathbb{R}^n$. Here, A_m^n stands for $A_m, \ldots, A_n$, and $\mathbb{R}$ denotes the set of real numbers. If there is a feedback link, then X_k, $k = 1, \ldots, n$, is a function of the message M and, additionally, of the past channel output symbols Y_1^{k-1}, i.e., $X_k = \varphi_n^{(k)}(M, Y_1^{k-1})$ for some mapping $\varphi_n^{(k)} : \mathcal{M} \times \mathbb{R}^{k-1} \to \mathbb{R}$. The receiver guesses the transmitted message M based on the n channel output symbols Y_1^n, i.e., $\hat{M} = \psi_n(Y_1^n)$ for some mapping $\psi_n : \mathbb{R}^n \to \mathcal{M}$.

Let $\mathbb{Z}^+$ denote the set of positive integers. The channel output $Y_k \in \mathbb{R}$ at time $k \in \mathbb{Z}^+$ corresponding to the channel inputs $(x_1, \ldots, x_k) \in \mathbb{R}^k$ is given by

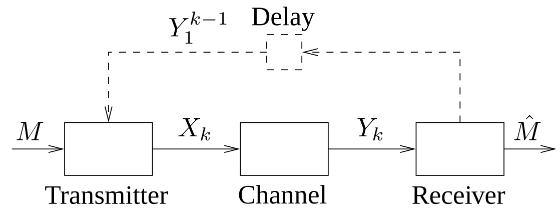

Fig. 1. The communication system.

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\nu=1}^{k-1} \alpha_{k-\nu} x_\nu^2\right)} \cdot U_k \tag{1}$$

where $\{U_k\}$ are independent and identically distributed (IID), zero-mean, unit-variance Gaussian random variables drawn independently of M. The coefficients $\{\alpha_\nu\}$ are non-negative and satisfy[1]

$$\sum_{\nu=1}^{\infty} \alpha_\nu \triangleq \alpha < \infty. \tag{2}$$

Note that this channel is not stationary as the variance of the additive noise depends on the time-index k.

We study the above channel under an average-power constraint on the inputs, i.e.,

$$\frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left[X_k^2\right] \le \mathrm{P}, \tag{3}$$

and we define the signal-to-noise ratio (SNR) as

$$\text{SNR} \triangleq \frac{\mathrm{P}}{\sigma^2}. \tag{4}$$

B. Capacity per Unit Cost

Let the *rate* R (in nats per channel use) be defined as

$$R \triangleq \frac{\log |\mathcal{M}|}{n} \tag{5}$$

where $\log(\cdot)$ denotes the natural logarithm function. A rate is said to be *achievable* if there exists a sequence of mappings ϕ_n (without feedback) or $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ (with feedback) and ψ_n such that the error probability $\Pr\left(\hat{M} \neq M\right)$ tends to zero as n goes to infinity. The *capacity* C is the supremum of all

[1] For on-chip communication the coefficients characterize the cool-down behavior of the chip, and it thus seems reasonable to assume that the coefficients are monotonically non-increasing, i.e., $\alpha_\nu \le \alpha_{\nu'}$ for $\nu \ge \nu'$. This assumption is, however, not required for the results stated in this paper.

achievable rates. We denote by $C(\mathrm{SNR})$ the capacity under the input constraint (3) when there is no feedback, and we add the subscript "FB" to indicate that there is a feedback link. Clearly,

$$C(\mathrm{SNR}) \leq C_{\mathrm{FB}}(\mathrm{SNR}) \tag{6}$$

as we can always ignore the feedback link.

In this paper we study the *capacities per unit cost* which are defined as [1]

$$\dot{C}(0) \triangleq \sup_{\mathrm{SNR}>0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \quad \text{and} \quad \dot{C}_{\mathrm{FB}}(0) \triangleq \sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}}. \tag{7}$$

Note that (6) implies

$$\dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0). \tag{8}$$

C. The Main Result

Our main result is stated in the following theorem.

Theorem 1: Consider the above channel model. Then, irrespective of whether feedback is available or not, the corresponding capacity per unit cost is given by

$$\dot{C}(0) = \dot{C}_{\mathrm{FB}}(0) = \frac{1}{2}(1+\alpha) \tag{9}$$

where α is defined in (2).

Theorem 1 is proved in Section II. In Section III we briefly discuss the above channel at high SNR. Specifically, we present a sufficient and a necessary condition on the coefficients $\{\alpha_\nu\}$ for capacity to be bounded in the SNR.

II. Proof of Theorem 1

In Section II-A we derive an upper bound on the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$, and in Section II-B we derive a lower bound on the capacity $C(\mathrm{SNR})$ in the absence of feedback. These bounds are then used in Section II-C to derive an upper bound on $\dot{C}_{\mathrm{FB}}(0)$ and a lower bound on $\dot{C}(0)$, and it is shown that both bounds are equal to $1/2 \cdot (1+\alpha)$. Together with (8) this proves Theorem 1.

A. Upper Bound

As in [2, Sec. 8.12], the upper bound on $C_{\mathrm{FB}}(\mathrm{SNR})$ is based on Fano's inequality and on an upper bound on $\frac{1}{n}I(M;Y_1^n)$, which for our channel can be expressed, using the chain rule for mutual information, as

$$\begin{aligned}
&\frac{1}{n}I(M;Y_1^n)\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[h(Y_k|Y_1^{k-1}) - h(Y_k|Y_1^{k-1},M)\right]\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[h(Y_k|Y_1^{k-1}) - h(Y_k|Y_1^{k-1},M,X_1^k)\right]\\
&= \frac{1}{n}\sum_{k=1}^{n}\Bigg[h(Y_k|Y_1^{k-1}) - h(U_k)\\
&\qquad\qquad - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\nu=1}^{k-1}\alpha_{k-\nu}X_\nu^2\right)\right]\Bigg]
\end{aligned} \tag{10}$$

where the second equality follows because X_1^k is a function of M and Y_1^{k-1}; and the last equality follows from the behavior of differential entropy under translation and scaling [2, Thms. 9.6.3 & 9.6.4], and because U_k is independent of (Y_1^{k-1}, M, X_1^k).

Evaluating the differential entropy $h(U_k)$ of a Gaussian random variable, and using the trivial lower bound $\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\nu=1}^{k-1}\alpha_{k-\nu}X_\nu^2\right)\right] \geq \log\sigma^2$, we obtain the final upper bound

$$\begin{aligned}
&\frac{1}{n}I(M;Y_1^n)\\
&\leq \frac{1}{n}\sum_{k=1}^{n}\left[h(Y_k|Y_1^{k-1}) - \frac{1}{2}\log(2\pi e\sigma^2)\right]\\
&\leq \frac{1}{2}\frac{1}{n}\sum_{k=1}^{n}\log\left(1 + \sum_{\nu=1}^{k}\alpha_{k-\nu}\mathsf{E}\left[X_\nu^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log\left(1 + \frac{1}{n}\sum_{k=1}^{n}\sum_{\nu=1}^{k}\alpha_{k-\nu}\mathsf{E}\left[X_\nu^2\right]/\sigma^2\right)\\
&= \frac{1}{2}\log\left(1 + \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\sum_{\nu=0}^{n-k}\alpha_\nu\right)\\
&\leq \frac{1}{2}\log\left(1 + (1+\alpha)\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log\left(1+(1+\alpha)\cdot\mathrm{SNR}\right)
\end{aligned} \tag{11}$$

where we define $\alpha_0 \triangleq 1$. Here, the second inequality follows because conditioning cannot increase entropy and from the entropy maximizing property of Gaussian random variables [2, Thm. 9.6.5]; the next inequality follows by Jensen's inequality; the following equality by rewriting the double sum; the subsequent inequality follows because the coefficients are nonnegative which implies that $\sum_{\nu=0}^{n-k}\alpha_\nu \leq \sum_{\nu=0}^{\infty}\alpha_\nu = 1+\alpha$; and the last inequality follows from the power constraint (3).

B. Lower Bound

As aforementioned, the above channel (1) is not stationary, and one therefore needs to exercise some care in relating the capacity $C(\mathrm{SNR})$ in the absence of feedback to the quantity

$$\varliminf_{n\to\infty}\frac{1}{n}\sup I(X_1^n;Y_1^n) \tag{12}$$

(where the maximization is over all input distributions satisfying the power constraint (3)). In fact, it is *prima facie* not clear whether there is a coding theorem associated with (12). We shall sidestep this problem by studying the capacity of a different channel whose time-k channel output $\tilde{Y}_k \in \mathbb{R}$ is, conditional on the sequence $\{X_k\} = \{x_k\}$, given by

$$\tilde{Y}_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\nu=-\infty}^{k-1}\alpha_{k-\nu}x_\nu^2\right)}\cdot U_k \tag{13}$$

where $\{U_k\}$ and $\{\alpha_\nu\}$ are defined in Section I-A. This channel has the advantage that it is stationary & ergodic in the sense

that when $\{X_k\}$ is a stationary & ergodic process then the pair $\{(X_k, \tilde{Y}_k)\}$ is jointly stationary & ergodic. It follows that if the sequences $\{X_k\}_{k\leq 0}$ and $\{X_k\}_{k\geq 1}$ are independent of each other, and if the random variables X_k, $k \leq 0$, are bounded, then any rate that can be achieved over this new channel is also achievable over the original channel. Indeed, the original channel (1) can be converted into (13) by adding

$$S_k = \sqrt{\left(\sum_{\nu=-\infty}^{0} \alpha_{k-\nu} X_\nu^2\right)} \cdot U_{-k}$$

to the channel output Y_k, and, since the independence of $\{X_k\}_{k\leq 0}$ and $\{X_k\}_{k\geq 1}$ ensures that the sequence $\{S_k\}$ is independent of the message M, it follows that any rate achievable over (13) can be achieved over (1) by using a receiver that generates $\{S_k\}$ and guesses then M based on $\{Y_k + S_k\}_{k=1}^n$.[2]

We consider $\{X_k\}$ that are block-wise IID in blocks of L symbols. Thus, denoting $\mathbf{X}_\ell = (X_{\ell L+1}, \ldots, X_{(\ell+1)L})^\mathsf{T}$ (where $(\cdot)^\mathsf{T}$ denotes the transpose), $\{\mathbf{X}_\ell\}$ are IID with $\mathbf{X}_\ell$ taking on the value $(\xi, 0, \ldots, 0)^\mathsf{T}$ with probability δ and $(0, \ldots, 0)^\mathsf{T}$ with probability $1-\delta$, for some $\xi \in \mathbb{R}$. Note that to satisfy the average-power constraint (3) we shall choose ξ and δ so that

$$\frac{\xi^2}{\sigma^2}\delta = L \cdot \text{SNR}. \tag{14}$$

Let $\tilde{\mathbf{Y}}_\ell = (\tilde{Y}_{\ell L+1}, \ldots, \tilde{Y}_{(\ell+1)L})^\mathsf{T}$, and let $\lfloor \cdot \rfloor$ denote the floor function. Noting that the pair $\{(\mathbf{X}_\ell, \tilde{\mathbf{Y}}_\ell)\}$ is jointly stationary & ergodic, it follows from [3] that the rate

$$R = \lim_{n\to\infty} \frac{1}{n} I\left(\mathbf{X}_0^{\lfloor n/L \rfloor -1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor -1}\right) \tag{15}$$

is achievable over the new channel (13) and, thus, yields a lower bound on the capacity $C(\text{SNR})$ of the original channel (1). We lower bound $\frac{1}{n} I(\mathbf{X}_0^{\lfloor n/L \rfloor -1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor -1})$ as

$$\begin{aligned}
&\frac{1}{n} I\left(\mathbf{X}_0^{\lfloor n/L \rfloor -1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor -1}\right)\\
&= \frac{1}{n} \sum_{\ell=0}^{\lfloor n/L \rfloor -1} I\left(\mathbf{X}_\ell; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor -1} \middle| \mathbf{X}_0^{\ell-1}\right)\\
&\geq \frac{1}{n} \sum_{\ell=0}^{\lfloor n/L \rfloor -1} I\left(\mathbf{X}_\ell; \tilde{\mathbf{Y}}_\ell \middle| \mathbf{X}_0^{\ell-1}\right)\\
&\geq \frac{1}{n} \sum_{\ell=0}^{\lfloor n/L \rfloor -1} \left[I\left(\mathbf{X}_\ell; \tilde{\mathbf{Y}}_\ell \middle| \mathbf{X}_{-\infty}^{\ell-1}\right) - I\left(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_\ell \middle| \mathbf{X}_0^{\ell}\right)\right]
\end{aligned} \tag{16}$$

where we use the chain rule and that reducing observations cannot increase mutual information. By using that (2) implies

$$\lim_{\ell\to\infty} \sum_{\nu=\ell}^{\infty} \alpha_\nu = 0$$

[2]The boundedness of the random variables X_k, $k \leq 0$, guarantees that the quantity $\sum_{\nu=-\infty}^{0} \alpha_{k-\nu} x_\nu^2$ is finite for any realization of $\{X_k\}_{k\leq 0}$.

it can be shown that the second term in the sum on the right-hand side (RHS) of (16) vanishes as ℓ tends to infinity. This together with a Cesáro type theorem [2, Thm. 4.2.3] yields

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} I\left(\mathbf{X}_0^{\lfloor n/L \rfloor -1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L \rfloor -1}\right)\\
&\geq \frac{1}{L} I\left(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \middle| \mathbf{X}_{-\infty}^{-1}\right)\\
&\quad - \lim_{n\to\infty} \frac{1}{L} \frac{1}{\lfloor n/L \rfloor} \sum_{\ell=0}^{\lfloor n/L \rfloor -1} I\left(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_\ell \middle| \mathbf{X}_0^{\ell}\right)\\
&= \frac{1}{L} I\left(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \middle| \mathbf{X}_{-\infty}^{-1}\right)
\end{aligned} \tag{17}$$

where the first inequality follows by the stationarity of $\{(\mathbf{X}_\ell, \tilde{\mathbf{Y}}_\ell)\}$ which implies that $I(\mathbf{X}_\ell; \tilde{\mathbf{Y}}_\ell | \mathbf{X}_{-\infty}^{\ell-1})$ does not depend on ℓ, and by noting that, for a fixed L, $\lim_{n\to\infty} \frac{\lfloor n/L \rfloor L}{n} = 1$.

We proceed to analyze $I(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 | \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$ for a given sequence $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$. Making use of the canonical decomposition of mutual information (e.g., [1, eq. (10)]), we have

$$\begin{aligned}
&I\left(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \middle| \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\right)\\
&= I\left(X_1; \tilde{\mathbf{Y}}_0 \middle| \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\right)\\
&= \int D\left(f_{\tilde{\mathbf{Y}}_0|X_1=x,\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right) \mathrm{d}P_{X_1}(x)\\
&\quad - D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right)\\
&= \delta D\left(f_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right)\\
&\quad - D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right)
\end{aligned} \tag{18}$$

where the first equality follows because, for our choice of input distribution, $X_2 = \ldots = X_L = 0$ and, hence, X_1 conveys as much information about $\tilde{\mathbf{Y}}_0$ as $\mathbf{X}_0$. Here, $D(\cdot\|\cdot)$ denotes relative entropy, and $f_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}}$, $f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}$, and $f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}$ denote the densities of $\tilde{\mathbf{Y}}_0$ conditional on the inputs $\left(X_1 = \xi, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\right)$, $\left(X_1 = 0, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\right)$, and $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$, respectively. Thus, $f_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}}$ is the density of an L-variate Gaussian random vector of mean $(\xi, 0, \ldots, 0)^\mathsf{T}$ and of diagonal covariance matrix $\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}$ with diagonal entries

$$\begin{aligned}
\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}(1,1) &= \sigma^2 + \sum_{i=-\infty}^{-1} \alpha_{-iL} x_{iL+1}^2\\
\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(\xi)}(k,k) &= \sigma^2 + \alpha_{k-1}\xi^2 + \sum_{i=-\infty}^{-1} \alpha_{-iL+k-1} x_{iL+1}^2,\\
&\qquad\qquad k = 2, \ldots, L;
\end{aligned}$$

$f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}$ is the density of an L-variate, zero-mean Gaussian random vector of diagonal covariance matrix $\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(0)}$ with diagonal entries

$$\mathsf{K}_{\mathbf{x}_{-\infty}^{-1}}^{(0)}(k,k) = \sigma^2 + \sum_{i=-\infty}^{-1} \alpha_{-iL+k-1} x_{iL+1}^2, \quad k = 1, \ldots, L;$$

and $f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}$ is given by

$$f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} = \delta f_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}} + (1-\delta) f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}.$$

In order to evaluate the first term on the RHS of (18) we note that the relative entropy of two real, L-variate Gaussian random vectors of the respective means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and of the respective covariance matrices K_1 and K_2 is given by

$$\begin{aligned} &D\left(\mathcal{N}(\boldsymbol{\mu}_1,\mathsf{K}_1)\,\|\mathcal{N}(\boldsymbol{\mu}_2,\mathsf{K}_2)\right) \\ &= \frac{1}{2}\log\det\mathsf{K}_2 - \frac{1}{2}\log\det\mathsf{K}_1 + \frac{1}{2}\operatorname{tr}\left(\mathsf{K}_1\mathsf{K}_2^{-1} - \mathsf{I}_L\right) \\ &\quad + \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\mathsf{K}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \end{aligned} \tag{19}$$

with $\det\mathsf{A}$ and $\operatorname{tr}(\mathsf{A})$ denoting the determinant and the trace of the matrix A, respectively, and where I_L denotes the $L\times L$ identity matrix. The second term on the RHS of (18) is analyzed in the next subsection.

Let $\mathsf{E}\left[D(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\|f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}})\right]$ denote the second term on the RHS of (18) averaged over $\mathbf{X}_{-\infty}^{-1}$, i.e.,

$$\begin{aligned} &\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right] \\ &\qquad = \mathsf{E}_{\mathbf{X}_{-\infty}^{-1}}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right)\right]. \end{aligned}$$

Then, using (19) & (18) and taking expectations over $\mathbf{X}_{-\infty}^{-1}$ we obtain, again defining $\alpha_0 \triangleq 1$,

$$\begin{aligned} &\frac{1}{L}I(\mathbf{X}_0;\tilde{\mathbf{Y}}_0\,|\mathbf{X}_{-\infty}^{-1}) \\ &= \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{k=1}^{L}\mathsf{E}\left[\frac{\alpha_{k-1}}{1+\sum_{i=-\infty}^{-1}\alpha_{-iL+k-1}X_{iL+1}^2/\sigma^2}\right] \\ &\quad - \frac{\delta}{L}\frac{1}{2}\sum_{k=2}^{L}\mathsf{E}\left[\log\left(1+\frac{\alpha_{k-1}\xi^2}{\sigma^2+\sum_{i=-\infty}^{-1}\alpha_{-iL+k-1}X_{iL+1}^2}\right)\right] \\ &\quad - \frac{1}{L}\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right] \\ &\geq \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{k=1}^{L}\frac{\alpha_{k-1}}{1+\sum_{i=-\infty}^{-1}\alpha_{-iL+k-1}\mathsf{E}\left[X_{iL+1}^2\right]/\sigma^2} \\ &\quad - \frac{\delta}{L}\frac{1}{2}\sum_{k=2}^{L}\log\left(1+\alpha_{k-1}\xi^2/\sigma^2\right) \\ &\quad - \frac{1}{L}\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right] \\ &\geq \frac{1}{2}\mathrm{SNR}\sum_{k=1}^{L}\frac{\alpha_{k-1}}{1+\alpha\cdot L\cdot\mathrm{SNR}} \\ &\quad - \frac{1}{2}\mathrm{SNR}\sum_{k=2}^{L}\frac{\log\left(1+\alpha_{k-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\ &\quad - \frac{1}{L}\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right] \end{aligned} \tag{20}$$

where the first inequality follows by the lower bound $\mathsf{E}[1/(1+X)] \geq 1/(1+\mathsf{E}[X])$ which is a consequence of Jensen's inequality applied to the convex function $1/(1+x)$, $x>0$, and by the upper bound

$$\begin{aligned} &\mathsf{E}\left[\log\left(1+\frac{\alpha_{k-1}\xi^2}{\sigma^2+\sum_{i=-\infty}^{-1}\alpha_{-iL+k-1}X_{iL+1}^2}\right)\right] \\ &\qquad\qquad \leq \log\left(1+\alpha_{k-1}\xi^2/\sigma^2\right) \end{aligned}$$

for every $k=2,\ldots,L$; and the second inequality follows by (14) and by upper bounding

$$\sum_{i=-\infty}^{-1}\alpha_{-iL+k-1} \leq \sum_{i=1}^{\infty}\alpha_i = \alpha$$

for every $k=1,\ldots,L$.

The final lower bound follows now by (20) and (17)

$$\begin{aligned} &\lim_{n\to\infty}\frac{1}{n}I\left(\mathbf{X}_0^{\lfloor n/L\rfloor-1};\tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\right) \\ &\geq \frac{1}{2}\mathrm{SNR}\sum_{k=1}^{L}\frac{\alpha_{k-1}}{1+\alpha\cdot L\cdot\mathrm{SNR}} \\ &\quad - \frac{1}{2}\mathrm{SNR}\sum_{k=2}^{L}\frac{\log\left(1+\alpha_{k-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\ &\quad - \frac{1}{L}\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right]. \end{aligned} \tag{21}$$

C. Asymptotic Analysis

We start with analyzing the upper bound (11). We have

$$\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{\frac{1}{2}\log(1+(1+\alpha)\cdot\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha) \tag{22}$$

where the second inequality follows by upper bounding $\log(1+x)\leq x$, $x>0$, and we thus obtain

$$\dot{C}_{\mathrm{FB}}(0) = \sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha). \tag{23}$$

In order to derive a lower bound on $\dot{C}(0)$ we first note that

$$\dot{C}(0) = \sup_{\mathrm{SNR}>0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \geq \lim_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \tag{24}$$

and proceed by analyzing the limiting ratio of the lower bound (21) to the SNR as the SNR tends to zero.

To this end, we first shall show that

$$\lim_{\mathrm{SNR}\downarrow 0}\frac{\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\right)\right]}{\mathrm{SNR}} = 0. \tag{25}$$

It was shown in [1, p. 1023] that for any pair of densities f_0 and f_1 satisfying $D(f_1\|f_0)<\infty$

$$\lim_{\beta\downarrow 0}\frac{D\left(\beta f_1+(1-\beta)f_0\|\, f_0\right)}{\beta} = 0. \tag{26}$$

Thus, for any given $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$, (26) together with $\delta = \mathrm{SNR}\cdot L\cdot\sigma^2/\xi^2$ implies that

$$\lim_{\mathrm{SNR}\downarrow 0}\frac{D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\right)}{\mathrm{SNR}} = 0. \tag{27}$$

In order to show that this also holds when $D(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}})$ is averaged over $\mathbf{X}_{-\infty}^{-1}$, we derive in the following the uniform upper bound

$$\sup_{\mathbf{x}_{-\infty}^{-1}} D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right) = D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right)\Big|_{\mathbf{x}_{-\infty}^{-1}=0}. \tag{28}$$

The claim (25) follows then by upper bounding

$$\mathsf{E}\left[D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}} \right) \right] \leq D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right)\Big|_{\mathbf{x}_{-\infty}^{-1}=0} \tag{29}$$

and by (27).

In order to prove (28) we use that any Gaussian random vector can be expressed as the sum of two independent Gaussian random vectors to write the channel output $\tilde{\mathbf{Y}}_0$ as

$$\tilde{\mathbf{Y}}_0 = \mathbf{X}_0 + \mathbf{V} + \mathbf{W} \tag{30}$$

where, conditional on $\mathbf{X}_{-\infty}^{0} = \mathbf{x}_{-\infty}^{0}$, $\mathbf{V}$ and $\mathbf{W}$ are L-variate, zero-mean Gaussian random vectors, drawn independently of each other, and having the respective diagonal covariance matrices $\mathsf{K}_{\mathbf{V}|\mathbf{x}_0}$ and $\mathsf{K}_{\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}$ whose diagonal entries are given by

$$\begin{aligned} \mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(1,1) &= \sigma^2 \\ \mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(k,k) &= \sigma^2 + \alpha_{k-1} x_1, \qquad k = 2, \ldots, L, \end{aligned}$$

and

$$\mathsf{K}_{\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}(k,k) = \sum_{i=-\infty}^{-1} \alpha_{-iL+k-1} x_{iL+1}^2, \quad k = 1, \ldots, L.$$

Thus, $\mathbf{V}$ is the portion of the noise due to $\mathbf{X}_0$, and $\mathbf{W}$ is the portion of the noise due to $\mathbf{X}_{-\infty}^{-1}$. Note that $\mathbf{X}_0 + \mathbf{V}$ and $\mathbf{W}$ are independent of each other because $\mathbf{X}_0$ is, by construction, independent of $\mathbf{X}_{-\infty}^{-1}$.

The upper bound (28) follows now by

$$\begin{aligned} & D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right) \\ &= D\left(f_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right) \\ &\leq D\left(f_{\mathbf{X}_0+\mathbf{V}} \| f_{\mathbf{X}_0+\mathbf{V}|X_1=0} \right) \\ &= D\left(f_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \middle\| f_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}} \right)\Big|_{\mathbf{x}_{-\infty}^{-1}=0} \end{aligned} \tag{31}$$

where $f_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}$ and $f_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}}$ denote the densities of $\mathbf{X}_0 + \mathbf{V} + \mathbf{W}$ conditional on the inputs $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$ and $(X_1 = 0, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$, respectively; $f_{\mathbf{X}_0+\mathbf{V}}$ denotes the unconditional density of $\mathbf{X}_0+\mathbf{V}$; and $f_{\mathbf{X}_0+\mathbf{V}|X_1=0}$ denotes the density of $\mathbf{X}_0 + \mathbf{V}$ conditional on $X_1 = 0$. Here, the inequality follows by the data processing inequality for relative entropy (see [2, Sec. 2.9]) and by noting that $\mathbf{X}_0 + \mathbf{V}$ is independent of $\mathbf{X}_{-\infty}^{-1}$.

Returning to the analysis of (21), we obtain from (24) and (25)

$$\begin{aligned} & \dot{C}(0) \\ & \geq \lim_{\mathrm{SNR}\downarrow 0} \frac{1}{2} \sum_{k=1}^{L} \frac{\alpha_{k-1}}{1 + \alpha \cdot L \cdot \mathrm{SNR}} - \frac{1}{2} \sum_{k=2}^{L} \frac{\log\left(1 + \alpha_{k-1} \frac{\xi^2}{\sigma^2}\right)}{\xi^2/\sigma^2} \\ & = \frac{1}{2} \sum_{k=1}^{L} \alpha_{k-1} - \frac{1}{2} \sum_{k=2}^{L} \frac{\log\left(1 + \alpha_{k-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}. \end{aligned} \tag{32}$$

By letting first ξ^2 go to infinity while holding L fixed, and by letting then L go to infinity, we obtain the desired lower bound on the capacity per unit cost

$$\dot{C}(0) \geq \frac{1}{2}(1 + \alpha). \tag{33}$$

Thus, (33), (8), and (23) yield

$$\frac{1}{2}(1 + \alpha) \leq \dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0) \leq \frac{1}{2}(1 + \alpha) \tag{34}$$

which proves Theorem 1.

III. High SNR Results

The channel described in Section I-A was studied at high SNR in [4] where it was asked whether capacity is bounded or unbounded in the SNR. It was shown that the answer to this question depends highly on the decay rate of the coefficients $\{\alpha_\nu\}$. We summarize the main result of [4] in the next theorem. For a statement of this theorem in its full generality and for a proof thereof we refer to [4].

Theorem 2: Consider the channel model described in Section I-A. Then,

$$\text{i)} \quad \varliminf_{\nu\to\infty} \frac{\alpha_{\nu+1}}{\alpha_\nu} > 0 \quad \Longrightarrow \quad \sup_{\mathrm{SNR}>0} C_{\mathrm{FB}}(\mathrm{SNR}) < \infty, \tag{35}$$

$$\text{ii)} \quad \varlimsup_{\nu\to\infty} \frac{\alpha_{\nu+1}}{\alpha_\nu} = 0 \quad \Longrightarrow \quad \sup_{\mathrm{SNR}>0} C(\mathrm{SNR}) = \infty, \tag{36}$$

where we define, for any $a > 0$, $a/0 \triangleq \infty$ and $0/0 \triangleq 0$.

For example, when $\{\alpha_\nu\}$ is a geometric sequence, i.e., $\alpha_\nu = \rho^\nu$ for $0 < \rho < 1$, then the capacity is bounded. Note that when neither the left-hand side (LHS) of (35) nor the LHS of (36) holds, i.e., when $\varlimsup_{\nu\to\infty} \alpha_{\nu+1}/\alpha_\nu > 0$ and $\varliminf_{\nu\to\infty} \alpha_{\nu+1}/\alpha_\nu = 0$, then the capacity can be bounded or unbounded.

Acknowledgment

Fruitful discussions with Ashish Khisti are gratefully acknowledged.

References

[1] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
[3] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
[4] T. Koch, A. Lapidoth, and P. P. Sotiriadis, "A hot channel," 2007, subm. to *IEEE Inform. Theory Workshop (ITW)*, Lake Tahoe, CA, USA. [Online]. Available: http://arxiv.org/abs/cs/0703154